RESEARCH ARTICLE

Estimating occupancy from autonomous recording unit data in the presence of misclassifications and detection heterogeneity

Matthew J. Clement[1] | **J. Andrew Royle**[2] | **Ronald J. Mixan**[1]

[1]Arizona Game and Fish Department, Phoenix, AZ, USA

[2]U.S. Geological Survey, Eastern Ecological Science Center, Laurel, MD, USA

Correspondence
Matthew J. Clement
Email: mclement@azgfd.gov

Abstract

1. Autonomous recording units (ARUs) are now widely used to survey communities of species. These surveys generate spatially and temporally replicated counts of unmarked animals, but such data typically include false negatives and misclassified detections, both of which may vary across sites in proportion to abundance. These data challenges can bias estimates of occupancy, and the typical approach of verifying individual detections is expensive.

2. We developed a Bayesian implementation of a two-species, false-positive N-mixture model for estimating occupancy from ARU data or other counts of unmarked animals that does not require manual verification. The model accounts for species misclassification and abundance-induced detection heterogeneity, as well as false negatives. To evaluate this model, we simulated 200 datasets for each of 29 scenarios, including scenarios in which misclassifications outnumbered correct classifications for rare species. We also applied the model to acoustic surveys of bats conducted on Fort Carson Army Post and Piñon Canyon Maneuver Site, Colorado, USA.

3. In the simulation study, bias, coverage and root mean square error for occupancy estimates obtained from the two-species false-positive N-mixture model were superior to metrics obtained from two competing two-species false-positive occupancy models. Across 29 scenarios, absolute bias was consistently low (range: −0.03 to 0.07), while coverage averaged 93% (range: 74%–98%). For alternative occupancy models, absolute bias was often high (range: −0.36 to 0.39), and coverage averaged from 47% to 65%. Although our model included an abundance parameter, abundance estimates were not reliable. For two species of *Myotis* bats, we estimated that 1%–5% of field-recorded detections were misclassified. Estimated occupancy (0.91 and 0.76) was lower than naïve estimates (1.00 and 0.94). Competing occupancy models implausibly estimated local occupancy of 0.00 at sites with numerous detections.

4. Our two-species, false-positive N-mixture model is significant because it accounts for detection heterogeneity and improves occupancy estimates without expensive manual verification of detections. Our field application indicated that misclassifications were not common, yet affected occupancy inferences. Given

that ARUs are increasingly used to survey a broad range of taxa, such an occupancy model could be widely useful.

1 | INTRODUCTION

Autonomous recording units (ARUs), such as acoustic recording devices or camera traps, have become important tools for monitoring wildlife (Gibb et al., 2019). Acoustic detectors are now widely used to monitor birds (Shonfield & Bayne, 2017), anurans (Chambert, Waddle, et al., 2018), cetaceans (Van Parijs et al., 2009) and fish (Luczkovich et al., 2008) while camera traps are used to monitor many large- and medium-sized terrestrial animals (O'Connell et al., 2010) as well as fish (Campbell et al., 2015). ARUs have been widely adopted because they can monitor many species simultaneously, their ability to run autonomously reduces field labour costs, they are minimally invasive, in some cases they can detect species that are resistant to other survey methods, and they create a digital record for future reference (Klingbeil & Willig, 2015). ARUs typically generate many detections, which are commonly classified using machine learning algorithms (Bittle & Duncan, 2013). Using such algorithms reduces data analysis costs and improves objectivity and repeatability of analysis (Clement, Murray, et al., 2014).

While ARUs and associated machine learning algorithms are powerful tools, they may generate detection errors, which may adversely affect inferences about species distributions (Miller et al., 2015). In addition to the false negatives that occur when a species is not detected, classification algorithms may misclassify a detection, resulting in a false-positive detection. Given that ARUs can record many detections, even a very low misclassification rate could generate many false positives. In fact, when inferences are desired for a rare species sympatric with common species, false-positive detections could be more numerous than true detections.

Occupancy models have been proposed to account for false-positive detections (Chambert et al., 2015; Miller et al., 2011; Royle & Link, 2006). However, false-positive occupancy models may have multiple solutions leading to ambiguous results (Royle & Link, 2006), which has typically been addressed with auxiliary data: either by collecting independent data that do not admit false positives (Miller et al., 2011) or by verifying classifications with a more reliable classification method (Chambert et al., 2015). In the context of acoustic surveys for bats, this often means that an expert will manually review species classifications to eliminate misclassifications (Chambert, Waddle, et al., 2018; Wright et al., 2020). However, expert review of bat acoustic recordings is time-consuming and there is ample evidence that it does not eliminate misclassifications (Barclay, 1999; Fritsch & Bruckner, 2014; Russo et al., 2018; Russo & Voigt, 2016).

Although false-positive occupancy models can address misclassified detections, they are vulnerable to detection heterogeneity, in which detection probability varies among sites (Clement, 2016; Miller et al., 2015). When unaccounted for, detection heterogeneity violates occupancy model assumptions and biases occupancy estimates (Royle & Nichols, 2003). Such bias corrupts hypothesis tests and may misdirect conservation actions. Although good study design and detection covariates can reduce detection heterogeneity (Miller et al., 2015), variation in abundance across sites is a pernicious source of detection heterogeneity (Royle & Nichols, 2003). While spatial variation in abundance is a universal truth in ecology, we note that spatial variation in acoustic detections of bats has been repeatedly documented, suggesting that occupancy analysis should account for detection heterogeneity (Baerwald & Barclay, 2009; Beilke et al., 2021; Gannon et al., 2003; Hayes, 1997; Sherwin et al., 2000).

We claim that ARU data contain information that could be used to address false-positive detections and abundance-induced detection heterogeneity without expensive and potentially fallible manual verification of detections. First, ARUs typically provide data on the number of detections at survey sites. Assuming classification algorithms perform better than random classification, sites where many recordings of a species of interest are recorded are more likely to represent true positives. Second, ARUs typically provide data for multiple species at survey sites. Because misclassifications of other species are likely to be the primary source of misclassifications (Miller et al., 2012), false-positive detections of a species of interest are more likely where many detections of other species have been recorded. It should be possible to obtain inferences about false-positive detections and occupancy probabilities by analysing the ensemble of detections with an appropriate probabilistic model, such as a multi-species false-positive occupancy model (Chambert, Waddle, et al., 2018; Wright et al., 2020). Furthermore, we expect to record more detections where a species is abundant. Assuming a positive relationship between abundance and detections, it is possible to account for detection heterogeneity by allowing the number of detections to vary with abundance (Royle, 2004; Royle & Nichols, 2003).

We propose that a two-species false-positive N-mixture model could account for misclassifications and abundance-induced detection heterogeneity when estimating occupancy from ARU data without manual verification of detections. Given the increased use of ARUs for various taxa, such a model could be useful across numerous applications. For example, the Arizona Game and Fish Department has

conducted acoustic monitoring for bats at Fort Carson Army Post and Piñon Canyon Maneuver Site in Colorado to assess the distribution of bats. While the recorded echolocations can be classified to species using commercial software, these data likely include misclassifications (Clement, Rodhouse, et al., 2014), making it difficult to estimate occupancy. Therefore, our objectives were to develop a two-species false-positive N-mixture model for use with ARUs and other types of replicated count data, test the model via simulations and apply the model to acoustic survey data for bats.

2 | MATERIALS AND METHODS

2.1 | Sampling design

We consider a sampling design commonly seen in ARU surveys. Within a 'season' (i.e. a brief time period that excludes significant mortality, recruitment or migration), detections of target species generated by ARUs are counted during $j = 1, 2, ..., T$ distinct sampling occasions at $i = 1, 2, ..., R$ sites. Such sampling results in counts, c_{ij}^A, for species A at site i during occasion j. We note that similar counts could arise by other detection methods, such as aural point counts, visual counts of groups of animals or machine learning analysis of images or video. These counts may include misclassification errors, in which the species is erroneously classified. We assume count data are collected for multiple species, such as during acoustic surveys for bats or camera trapping of medium or large mammals.

2.2 | Model description

We first provide context for our approach by reviewing two existing approaches to modelling occupancy when false-positive detections occur. First, consider a two-species false-positive occupancy model (Chambert, Grant, et al., 2018). As in other occupancy models, the presence of each species is described by a Bernoulli process:

$$z_i^A \sim \text{Bern}(\psi^A),$$

$$z_i^B \sim \text{Bern}(\psi^B), \tag{1}$$

where, for species A and B, z_i is the site-specific presence (1) or absence (0), and ψ is the unconditional probability of occupancy.

Detections for each species can also be described by Bernoulli processes. Because detections of one species may be incorrectly classified as a different species (Miller et al., 2012), the detection model allows false-positive detections when species B is misclassified as species A and vice versa:

$$y_{ij}^A \sim \text{Bern}(z_i^A p_{11}^A + (1 - z_i^A) z_i^B p_{10}^B),$$

$$y_{ij}^B \sim \text{Bern}(z_i^B p_{11}^B + (1 - z_i^B) z_i^A p_{10}^A), \tag{2}$$

where y_{ij}^A is the detection (1) or non-detection (0) of species A at site i during occasion j, p_{11}^A is the probability of detecting and correctly classifying species A given it is present, and p_{10}^A is the probability of detecting and misclassifying species A as species B, given species A is present and species B is absent, with equivalent parameters for species B. For each species, we also included the constraint that $p_{11}^A > p_{10}^A$ (i.e. species A is more likely to be correctly classified than misclassified), which is plausible and improves model identifiability (Royle & Link, 2006). Hereafter, we refer to the model described by Equations 1 and 2 as the Bernoulli detection model.

Second, we describe a two-species false-positive occupancy model that accounts for the number of detections recorded. For many types of ARU data, occupied sites may generate multiple, discrete detections. In this case, one option is to model the detection frequency as a Poisson count:

$$c_{ij} \sim \text{Poisson}(z_i \lambda),$$

where c_{ij} is the number of detections at site i during occasion j, and λ is an encounter rate parameter or intensity parameter expressing the average number of detections, given presence. This detection model can be modified to incorporate a false-positive encounter rate, ω (Chambert, Waddle, et al., 2018):

$$c_{ij} \sim \text{Poisson}(z_i \lambda + \omega).$$

In this formulation, the false-positive encounter rate is exogenous, occurring with the same intensity at all sites although, in principle, it could vary in response to covariates. Wright et al. (2020) proposed combining the two-species false-positive model of Chambert, Grant, et al. (2018) and the false-positive encounter rate model of Chambert, Waddle, et al. (2018) to generate a two-species false-positive encounter rate detection model:

$$c_{ij}^A \sim \text{Poisson}(z_i^A \lambda^A + z_i^B \omega^B),$$

$$c_{ij}^B \sim \text{Poisson}(z_i^B \lambda^B + z_i^A \omega^A), \tag{3}$$

where ω^A is a false-positive encounter rate given species A is present and all other parameters are defined above. We included, for each species, the constraint that $\lambda^A > \omega^A$ (i.e. the species classifier performs better than random classification), which is plausible and improves model identifiability analogous to the ordinary false-positive occupancy model (Royle & Link, 2006). Hereafter, we refer to the model described by Equations 1 and 3 as the Poisson detection model.

While the Bernoulli detection and Poisson detection models assume that detection or the encounter rate are constant across sites, detection may vary positively with abundance across sites, which can bias estimates of occupancy (Royle & Nichols, 2003). To address this bias, we propose to incorporate animal abundance as a latent variable in both the process and detection models. As such, our approach is a type of N-mixture model (Royle, 2004). Under a conventional N-mixture model, animals are distributed across a number of sites, making a Poisson or negative binomial distribution a natural choice for the

process model (Royle, 2004). We used the negative binomial distribution because it is suited to heterogeneous landscapes:

$$N_i^A \sim \text{NegBin}(\mu^A, r^A),$$

$$N_i^B \sim \text{NegBin}(\mu^B, r^B), \tag{4}$$

where, for species A and B, N_i is the site-specific abundance, μ is mean abundance and the over-dispersion parameter, r, is inversely related to variance so that $\text{Var}(N_i) = \mu + \mu^2/r$. Zero-truncated distributions or random effects could also be considered. The probability that $N_i > 0$ is equivalent to the usual occupancy parameter, ψ. For the negative binomial distribution, these parameters are related according to $\psi = 1 - \left(\frac{r}{r+\mu}\right)^r$.

To account for misclassifications, we propose a detection model similar to that in the Poisson detection model (Equation 3), but with detections proportional to abundance:

$$c_{ij}^A \sim \text{Poisson}(N_i^A \lambda^A + N_i^B \omega^B),$$

$$c_{ij}^B \sim \text{Poisson}(N_i^B \lambda^B + N_i^A \omega^A), \tag{5}$$

where all parameters have been defined above, but λ and ω now represent the expected number of true or false detections per individual, rather than detections per occupied site. We again included, for each species, the constraint that $\lambda^A > \omega^A$. Under this formulation, the count of detections depends directly on abundance, which is key to accounting for abundance-induced heterogeneity. Hereafter, we refer to the model in Equations 4 and 5 as the two-species false-positive N-mixture model, or briefly, the N-mixture model.

We note that the N-mixture model extends easily, in principle, to >2 species, where each new species contributes new encounter rate parameters to the Poisson intensity function. We propose several other extensions that could prove useful. For example, an exogenous source of misclassifications, ω^e, such as wind or 'imagined' animals (Alldredge et al., 2008) could be added to the count portion of the likelihood:

$$c_{ij}^A \sim \text{Poisson}(N_i^A \lambda^A + N_i^B \omega^B + \omega^{eA}) \text{ and } \lambda^A < \omega^A,$$

$$c_{ij}^B \sim \text{Poisson}(N_i^B \lambda^B + N_i^A \omega^A + \omega^{eB}) \text{ and } \lambda^B < \omega^B.$$

If survey data do not exist for multiple species, a single-species model could be specified in which all misclassifications are exogenous:

$$c_{ij}^A \sim \text{Poisson}(N_i^A \lambda^A + \omega^e).$$

We expect that in some circumstances, individual observations could be validated. For example, if scats of different species were counted, some or all observations during a subset of surveys could be validated using DNA analyses (Chambert et al., 2015). If validation occurs at m sites that are a subset of the R sites and during n occasions that are a subset of the T occasions then the two elements of the counts (true encounters and false-positive encounters) can be fully specified for a subset of surveys:

$$c_{ij}^A \sim \text{Poisson}(N_i^A \lambda^A + N_i^B \omega^B + \omega^{eA}) \text{ and } \lambda^A < \omega^A,$$

$$c_{ij}^B \sim \text{Poisson}(N_i^B \lambda^B + N_i^A \omega^A + \omega^{eB}) \text{ and } \lambda^B < \omega^B,$$

$$c_{mn}^A = c(\text{t})_{mn}^A + c(\text{f})_{mn}^A,$$

$$c(\text{t})_{mn}^A \sim \text{Poisson}(N_m^A \lambda^A),$$

$$c(\text{f})_{mn}^A \sim \text{Poisson}(N_m^B \lambda^B),$$

$$c_{mn}^B = c(\text{t})_{mn}^B + c(\text{f})_{mn}^B,$$

$$c(\text{t})_{mn}^B \sim \text{Poisson}(N_m^B \lambda^B),$$

$$c(\text{f})_{mn}^B \sim \text{Poisson}(N_m^A \lambda^A),$$

where $c(\text{t})$ is the count of true encounters and $c(\text{f})$ is the count of false encounters.

Alternatively, ARU data could be combined with 'certain' survey data, that is, data that exclude false positives (Miller et al., 2011). These data provide direct information about some model parameters or latent variables (analogous to 'supervised learning', while the models described by Equations 1–5 are a form of 'unsupervised learning'). In the context of bat surveys, capture surveys are a common survey method that avoids both double-counting and misclassifications. When these surveys are conducted during $k = 1$, 2, ..., S, occasions, they generate additional count data, x_{ik}, which can reasonably be modelled as a binomial process due to the lack of misclassifications:

$$x_{ik} \sim \text{Binomial}(q, N_i),$$

where q is the detection probability during certain surveys. Accordingly, Equation 5 could be augmented with certain surveys by adding the following terms:

$$x_{ik}^A \sim \text{Binomial}(q^A, N_i^A),$$

$$x_{ik}^B \sim \text{Binomial}(q^B, N_i^B).$$

2.3 | Simulation study

We performed a simulation study to verify that the two-species false-positive N-mixture model is identifiable, and to compare its performance to alternative models. We simulated data under the proposed two-species false-positive N-mixture model using a range of input parameters. In all simulations, $T = 8$ surveys were performed at $R = 60$ sites. We varied (a) the mean site abundance (by adjusting μ^A and μ^B in tandem), (b) the variance in abundance across sites (by adjusting the over-dispersion parameters r^A and r^B in tandem), (c) the overall encounter rate (by adjusting the true-positive, λ^A and λ^B, and false-positive, ω^A and ω^B, encounter rates in tandem), (d) the

probability a signal will be misclassified (by adjusting ω^A and ω^B in tandem), (e) the relative abundance of the two species (by adjusting μ^B independently of μ^A), (f) the relative signal production of the two species (by adjusting λ^B and ω^B independently of λ^A and ω^A), and (g) extra-Poisson variation in signal production (by adding a temporal random effect term, having standard deviation σ, to data simulation). These combinations yielded 29 simulation scenarios (Table 1). The false-positive occupancy parameters, ψ, p_{11} and p_{10} were not directly set in the simulations, but were determined by μ, r, λ and ω.

In particular, within any one scenario, p_{11} and p_{10} varied among sites due to variation in N_i (Figure S1). Our scenarios included simulations in which the majority of detections of species A were actually false positives (Scenarios 20, 21, 23, 24 and 25), which we consider to be a difficult challenge for estimation, yet biologically realistic if the target species is rare.

We simulated data 200 times under each scenario in Program R (v 4.04, R Core Team, 2021). For each simulated dataset, we estimated parameters of interest using three models: the Bernoulli detection model

TABLE 1 Scenarios used for data simulation and analysis. Each simulation includes two species (A and B), $T = 8$ surveys, and $R = 60$ sites. Table shows parameters that varied among simulations, including μ (mean site abundance), r (over-dispersion parameter), λ (mean correctly classified detections per individual), ω (mean misclassified detections per individual), Var(μ)/μ (factor by which variance in detections exceeds Poisson variance; controlled by the random effect standard deviation, σ). Bold font highlights differences among scenarios.

Factor of interest	Scenario	μ^A	r^A	λ^A	ω^A	$\mu^A:\mu^B$	$\lambda^A:\lambda^B$	Var(μ^A)/μ^A
Abundance	1	**0.33**	0.5	1	0.09	1:1	1:1	1
	2	**1**	0.5	1	0.09	1:1	1:1	1
	3	**3**	0.5	1	0.09	1:1	1:1	1
	4	**9**	0.5	1	0.09	1:1	1:1	1
	5	**27**	0.5	1	0.09	1:1	1:1	1
Overdispersion	6	3	**0.125**	1	0.09	1:1	1:1	1
	3	3	**0.5**	1	0.09	1:1	1:1	1
	7	3	**2**	1	0.09	1:1	1:1	1
	8	3	**8**	1	0.09	1:1	1:1	1
	9	3	**32**	1	0.09	1:1	1:1	1
Detections per individual	10	3	0.5	**0.25**	**0.02**	1:1	1:1	1
	3	3	0.5	**1**	**0.09**	1:1	1:1	1
	11	3	0.5	**4**	**0.35**	1:1	1:1	1
	12	3	0.5	**16**	**1.39**	1:1	1:1	1
	13	3	0.5	**64**	**5.57**	1:1	1:1	1
Misclassifications	14	3	0.5	1	**0.02**	1:1	1:1	1
	15	3	0.5	1	**0.04**	1:1	1:1	1
	3	3	0.5	1	**0.09**	1:1	1:1	1
	16	3	0.5	1	**0.19**	1:1	1:1	1
	17	3	0.5	1	**0.47**	1:1	1:1	1
Relative abundance	1	0.33	0.5	1	0.09	**1:1**	1:1	1
	18	0.33	0.5	1	0.09	**1:3**	1:1	1
	19	0.33	0.5	1	0.09	**1:9**	1:1	1
	20	0.33	0.5	1	0.09	**1:27**	1:1	1
	21	0.33	0.5	1	0.09	**1:81**	1:1	1
Relative detections per individual	10	3	0.5	0.25	0.02	1:1	**1:1**	1
	22	3	0.5	0.25	0.02	1:1	**1:4**	1
	23	3	0.5	0.25	0.02	1:1	**1:16**	1
	24	3	0.5	0.25	0.02	1:1	**1:64**	1
	25	3	0.5	0.25	0.02	1:1	**1:256**	1
Extra-Poisson variance	3	3	0.5	1	0.09	1:1	1:1	**1**
	26	3	0.5	1	0.09	1:1	1:1	**4.0**
	27	3	0.5	1	0.09	1:1	1:1	**9.5**
	28	3	0.5	1	0.09	1:1	1:1	**25.7**
	29	3	0.5	1	0.09	1:1	1:1	**79.4**

(Chambert, Grant, et al., 2018; Equations 1 and 2), the Poisson detection model (Wright et al., 2020; Equations 1 and 3) and our two-species false-positive N-mixture model (Equations 4 and 5) which was the data-generating model (except when we added temporal random effects to the data in Scenarios 26–29). We wrote the likelihoods described above in the JAGS programming language and evaluated them in JAGS (v 4.3.0, Plummer, 2003) using the jagsUI (v 1.5.1, Kellner, 2019) interface in Program R. JAGS evaluates likelihoods via Markov chain Monte Carlo (MCMC) simulation. We used flat priors for all parameters, including a diffuse half-normal distribution for encounter rates (λ and ω), a diffuse gamma distribution for abundance parameters (μ and r), and a uniform distribution for probabilities (ψ, p_{11} and p_{10}). We generated parameter estimates from three chains per model, with up to 100,000 iterations, after discarding the initial 10,000 iterations. Thinning by a factor of 3 yielded up to 100,000 posterior estimates for each parameter. We inspected the Gelman–Rubin convergence statistic ($\hat{r}$, Gelman & Rubin, 1992) and retained all simulations in which the occupancy parameters and model deviance converged (all $\hat{r} < 1.10$).

We calculated convergence, bias, coverage and root mean square error for the estimated occupancy of species A. Convergence measures if the MCMC simulation generated stable estimates and was calculated as the per cent of simulations in which the Gelman–Rubin convergence statistic was <1.10 for occupancy parameters and model deviance. Bias measures if estimates are close to the true value and was calculated as the difference between the estimated parameter value and the true parameter value. Coverage measures if credible intervals accurately represent uncertainty in the data and was calculated as the per cent of simulations in which the 95% credible interval overlaps the expected parameter value. Root mean square error measures if estimates are consistent and was calculated as the square root of the mean of the squared difference between each estimate and the expected value.

2.4 | Application to bat surveys

In addition to simulations, we applied our estimator to field data and evaluated the resulting estimates. The Arizona Game and Fish Department, in partnership with the U.S. Department of Defense, conducted passive acoustic surveys for bats at 18 sites in Fort Carson Army Post and Piñon Canyon Maneuver Site in southern and central Colorado (Figure 1; Mixan et al., 2020). These sites exist at the transition from the central shortgrass prairie to the southern Rocky Mountains (Chapman et al., 2006). Elevations range from 1,340 m to 2,102 m. The lower elevation areas are typically grasslands while the higher elevations support pinyon–juniper woodlands.

Acoustic survey stations were monitored from October 2018 to December 2019. At each station, we mounted a microphone and solar panel to a 2 m pole. Echolocation was automatically recorded by SM4 bat detectors (Wildlife Acoustics Inc) stored in weather-proof boxes. For this analysis, we selected acoustic data from 1 to 15 June 2019. We selected a time period that was long enough to detect many bats, but short enough that the occupancy status of sites should not change. We excluded two detector-nights (0.7%) due to poor weather (e.g. rain, wind) or technical problems (e.g. dead batteries). We used Kaleidoscope Pro version 5.1.9 software with the Bats of North America 5.1.0 Auto ID Classifier (Wildlife Acoustics Inc) to identify the recorded calls to species. To accept a species classification, we required Kaleidoscope to detect at least four individual calls within a call sequence, with a matching ratio ≥0.90, the default settings. If two bat passes (i.e. a sequence of echolocation calls) for a single species were recorded during the same minute, we recorded this as a single detection, on the grounds that these passes are often not independent events (Miller, 2001). We did not manually review
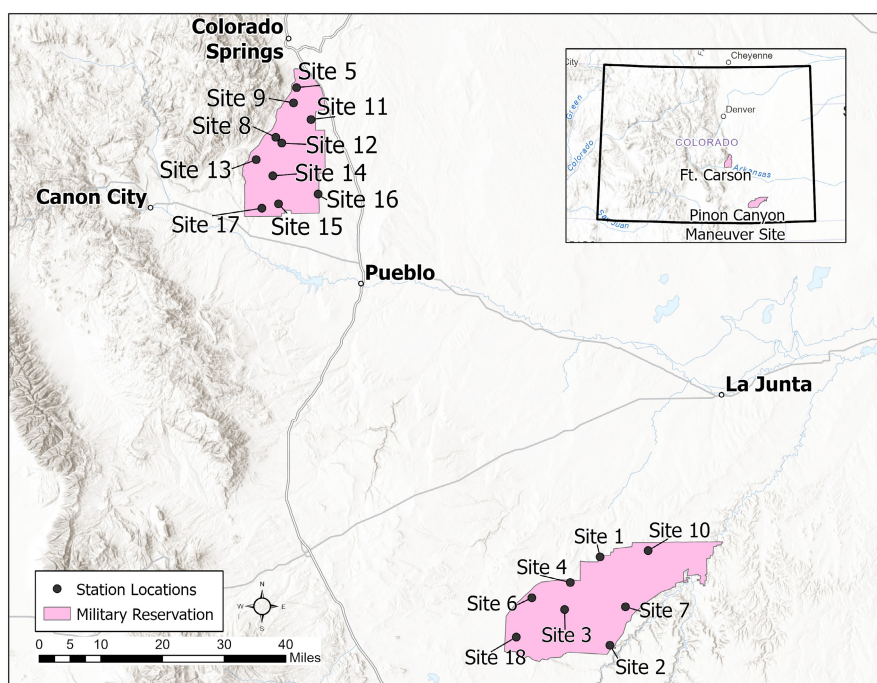

FIGURE 1 Locations of 18 passive acoustic monitoring stations at Fort Carson Army Post and Piñon Canyon Maneuver Site in Colorado, USA.

classifications due to the labour cost and subjectivity of a review. No animal capture permits were required for the acoustic surveys, as we did not capture any animals.

We selected two potentially confusing bat species for analysis with the three occupancy models (Bernoulli detection, Poisson detection and N-mixture model). Both *Myotis ciliolabrum* and *M. yumanensis* produce calls of 2 ms duration that rapidly descend in frequency to about 40 kHz (Thomas et al., 1987). For each occupancy model, we estimated the unconditional probability of occupancy (ψ), the conditional occupancy status at each recording site using the posterior mean ($\widetilde{z}_i = Pr(z_i = 1|\text{data})$; i.e. the probability the species was at a particular site, given the survey data), the probability of misclassifying a detection of each species $\left(\eta_A = \frac{\omega_A}{\lambda_A + \omega_A} \right)$, and the probability of a detection of a species being a false positive $\left(\tau_A = \frac{\omega_B}{\lambda_A + \omega_B} \right)$.

3 | RESULTS

3.1 | Simulation study

In the 29 simulation scenarios, the N-mixture performed best, with the lowest bias and root mean square error (RMSE), and the highest convergence and credible interval coverage, while the Poisson detection model performed poorly (Table S1). The N-mixture model exhibited low bias regardless of abundance, overdispersion of abundance, false-positive rate, an abundant non-target species, a highly detectable non-target species or unmodelled temporal random effects (Figure 2). Occupancy estimates for species *A* (the target species) were unbiased even when most detections of species *A* were false positives. The N-mixture model did exhibit some positive bias in occupancy when detections per animal ($\lambda^A + \omega^A$) were >16. The Bernoulli detection model achieved low bias regardless of abundance, false-positive rate or unmodelled temporal random effects, but was negatively biased by a large overdispersion parameter, many detections per animal, and a highly detectable non-target species, and it was positively biased by an abundant non-target species (Figure 2). The Poisson detection model often failed to produce estimates, with a convergence rate <50% in 25 of 29 scenarios (Table S1). When the Poisson detection model did converge, it was negatively biased in most scenarios, but positively biased by an abundant non-target species (Figure 2).

While the N-mixture model includes an abundance parameter, estimates of abundance were consistently biased high, and severely biased when detections per animal were high (Figure S2).

3.2 | Application to bat surveys

During 268 acoustic survey-nights, *M. ciliolabrum* was detected at all 18 sites, while *M. yumanensis* was detected at 17 (Table 2), yielding naïve occupancy rates of 1.00 and 0.94 respectively. For *M. ciliolabrum*, mean detections per night was 36.4, while the
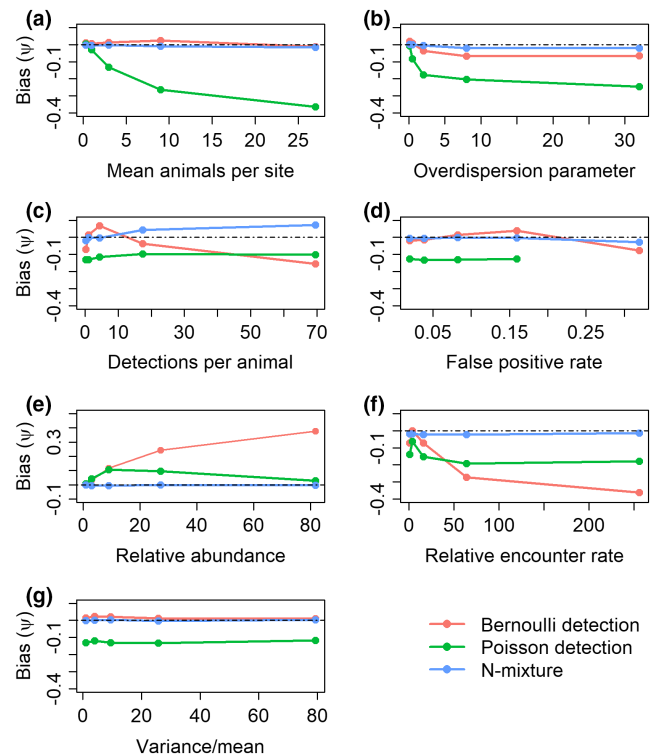

FIGURE 2 Bias in occupancy (ψ) estimates across 29 simulation scenarios for three occupancy models. Scenarios organized by (a) mean animals per site, (b) overdispersion parameter (note, a high parameter indicates low variance), (c) detections per animal, (d) false-positive rate (note, Poisson detection model did not converge at highest false-positive rate), (e) relative abundance (value indicates number of non-target animals per target animal), (f) relative encounter rate (value indicates ratio of signals produced per non-target animal to signals produced per target animal) and (g) extra-Poisson variance (value indicates factor by which variance exceeds Poisson variance). Black dotted line indicates zero bias.

variance across sites was 1,234.5, indicating high overdispersion and suggesting substantial detection heterogeneity. Two sites recorded fewer than 1 detection per night, raising the possibility that these detections were false positives. For *M. yumanensis*, considering the 17 sites with detections, overdispersion was even higher, with a mean detections per night of 31.5, and a variance across sites of 2727.2. Five sites recorded fewer than 1 detection per night, raising the possibility that these detections were false positives.

Estimates of the unconditional occupancy probability (ψ for the occupancy models and $1 - \left(\frac{r}{r+\mu} \right)^r$ for the N-mixture model) differed among estimators, although not significantly. The N-mixture model estimated the lowest rate of false positives and misclassification errors and the highest occupancy while the Poisson detection model estimated the highest rate of false positives and misclassification errors and the lowest occupancy (Table 2). While the unconditional occupancy probability estimates did not differ significantly among estimators, the site-specific estimates of conditional occupancy status did. Specifically, the N-mixture model estimated $\widetilde{z}_i > 0$ at every site for both species. N-mixture model estimates of $\widetilde{z}_i$ were

TABLE 2 Acoustic survey results and estimated parameters (and credible intervals) from three occupancy models applied to Arizona Game and Fish acoustic surveys of two species of bats: *Myotis ciliolabrum* (MYCI) and *M. yumanensis* (MYYU). The occupancy models are a two-species false-positive occupancy model with Bernoulli detection, a two-species false-positive occupancy model with Poisson detection and a two-species false-positive N-mixture model. Parameters include the unconditional probability of occupancy (ψ), probability of detection (p_{11}), probability of false-positive detection (p_{10}), average number of true detections (λ; per site for Poisson detection model and per individual for N-mixture model), average number of false detections (ω; per site for Poisson detection model and per individual for N-mixture model), probability that an actual detection of a species will be misclassified (η), the probability that a purported detection of a species is a false positive (τ) and the conditional probability that a given site is occupied given survey results ($\tilde{z}_i$). A '—' indicates not applicable.

Parameter	Detections/ detector-night MYCI	MYYU	Bernoulli detection model MYCI	MYYU	Poisson detection model MYCI	MYYU	N-mixture model MYCI	MYYU
ψ	—	—	0.82 (0.62–0.96)	0.73 (0.52–0.90)	0.70 (0.49–0.87)	0.53 (0.31–0.75)	0.91 (0.76–0.99)	0.76 (0.55–0.91)
p_{11}	—	—	0.84 (0.78–0.89)	0.75 (0.68–0.82)	—	—	—	—
p_{10}	—	—	0.06 (0.01–0.14)	0.10 (0.01–023)	—	—	—	—
λ	—	—	—	—	45.8 (42.9–47.9)	48.6 (45.0–54.5)	0.65 (0.13–1.22)	0.43 (0.13–0.82)
ω	—	—	—	—	5.28 (3.92–6.30)	6.63 (4.38–10.8)	0.01 (0.00–0.01)	0.02 (0.00–0.05)
η	—	—	—	—	0.10 (0.08–0.12)	0.12 (0.09–0.17)	0.01 (0.00–0.01)	0.05 (0.02–0.07)
τ	—	—	—	—	0.13 (0.09–0.20)	0.10 (0.07–012)	0.04 (0.01–0.14)	0.02 (0.00–0.04)
$\tilde{z}_1$	16.0	28.5	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_2$	4.3	70.7	0.99 (0.99–1.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.49 (0.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_3$	1.1	0.7	0.38 (0.00–1.00)	0.62 (0.00–1.00)	0.33 (0.00–1.00)	0.67 (0.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_4$	0.7	1.1	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_5$	48.3	38.6	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_6$	12.5	0.1	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.04 (0.00–1.00)
$\tilde{z}_7$	22.0	79.5	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.67 (0.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_8$	47.5	9.3	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_9$	105.0	4.6	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_{10}$	29.0	52.5	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_{11}$	46.5	13.2	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_{12}$	11.8	1.7	1.00 (1.00–1.00)	0.99 (0.99–1.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_{13}$	78.4	0.5	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.08 (0.00–1.00)
$\tilde{z}_{14}$	3.6	0.0	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.01 (0.00–0.01)
$\tilde{z}_{15}$	107.1	207.8	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_{16}$	52.5	0.7	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.35 (0.00–1.00)
$\tilde{z}_{17}$	69.6	26.5	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
$\tilde{z}_{18}$	0.1	0.2	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	1.00 (1.00–1.00)	0.22 (0.00–1.00)	0.99 (1.00–1.00)

low ($\tilde{z}_i < 0.1$) only when one species was undetected or detections were much higher for one species (Sites 6, 13, and 14). In contrast, the Poisson detection model estimated $\tilde{z}_i = 0$ at four sites for *M. ciliolabrum* and at eight sites for *M. yumanensis*. Of note, the Poisson detection model estimated $\tilde{z}_i = 0$ for *M. ciliolabrum* at Site 1 despite recording 16.0 bat passes per detector-night (and 28.5 bat passes per detector-night for *M. yumanensis*). The Bernoulli detection model estimated $\tilde{z}_i = 0$ at two sites for *M. ciliolabrum* and at four sites for *M. yumanensis*. Relatedly, the N-mixture model estimated that both species were present (i.e. $\tilde{z}_i > 0.5$ for both species) at 12 of 18 sites, while the Poisson detection model estimated that both species were present at just five sites, and the Bernoulli detection model estimated that both species were present at 11 sites.

4 | DISCUSSION

We developed a two-species false-positive N-mixture model that accounts for misclassifications in ARU data and allows for abundance-induced heterogeneity in encounter rates when estimating occupancy. The model is a blend of several existing models, including N-mixture models (Royle, 2004; Stanley & Royle, 2005) and occupancy models with features to account for false-positive detections (Royle & Link, 2006), multiple species (Chambert, Grant, et al., 2018; Wright et al., 2020) and Poisson encounter rates of species (Chambert, Waddle, et al., 2018; Wright et al., 2020). The model regards the abundance of each species as a latent variable, and thus factors which affect abundance can be modelled separately from those that affect detection

(Royle, 2004). We present the model as a Bayesian implementation of an extension to N-mixture models, which aids in modelling temporal or spatial variation in animal abundance or detection (Kéry & Royle, 2016). Across 29 scenarios, the N-mixture model outperformed two other false-positive occupancy models on simulated data, producing lower bias and RMSE, and higher coverage. It also produced more plausible estimates when applied to field data, as discussed below. In addition to superior statistical performance, we obtained our estimates without expensive validation of observations and we did not rely on the potentially invalid assumption that a portion of our data was error-free. Potential shortcomings of the N-mixture model are that it exhibited modest bias when individuals were detected multiple times, it was unable to reliably estimate abundance, and it may be sensitive to violations of model assumptions (Nakashima, 2020).

The occupancy models we compared to our N-mixture model were similar, but not identical to models found in the literature. Specifically, the Bernoulli detection model we tested differs from the two-species false-positive occupancy model by Chambert, Grant, et al. (2018), in that the latter model includes extensions in which the presence of one species affects the presence of the second species. In addition, Chambert, Grant, et al. (2018) used genetic analysis to confirm the identity of some individuals, while we did not use validation data in our model. Our Poisson detection model differs from the Wright et al. (2020) occupancy model in that their model used one total encounter rate parameter per species (Λ^A) and a vector of classification probabilities per species ($\theta_{AA}\ \theta_{AB}$) to decompose correctly and incorrectly classified encounters, while our detection parameters represent the products of their parameters ($\lambda^A = \Lambda^A\theta_{AA}$; $\omega^A = \Lambda^A\theta_{AB}$). In addition, Wright et al. (2020) used error-free manual confirmation of classifications to inform estimates of classification probabilities, while we did not. In preliminary testing, we found, as did Wright et al. (2020), that their model did not appear to be identifiable without confirmed observations. Our specification of the Poisson detection model (Equation 3) improved model convergence, to a degree, by eliminating one detection parameter per species (λ^A and ω^A vs. Λ^A, θ_{AA} and θ_{AB}). Wright et al. (2020) also proposed accounting for overdispersion in detections by replacing the Poisson distribution in Equation 3 with a log-normal distribution. In preliminary tests, we found that the log-normal detection model also did not converge without confirmed observations, so we did not test this model further.

After completing our analysis, we became aware that Clare et al. (2021) recently proposed a false-positive Royle–Nichols model. Such a model also includes abundance as a latent variable and may help account for abundance-induced detection heterogeneity (Royle & Nichols, 2003). It is difficult to directly compare their results to ours because they focused on predicting relative abundance, rather than occupancy, and they used manual verification of detections. Nonetheless, such a model could help account for detection heterogeneity when counts of detections are not available, such as during a frog chorus survey. Extending their model to multiple species might reduce the need for manual verification of detections, although this would need to be investigated.

We did not expect the poor performance of the Poisson detection model in our simulation study. Because the N-mixture model was the data-generating model, the Poisson detection model was misspecified, making some bias inevitable. However, we believe the data-generating model was a reasonable facsimile of biological data, making the severe bias of the Poisson detection model concerning. With hindsight, the bias appears to arise because the Poisson distribution predicts little variation in the number of detections across occupied sites. For example, if the mean number of detections is 52 (see scenario 12 in Table 1), the probability of obtaining 35 or fewer detections is just 0.8%. Therefore, depending on the number of detections observed for the other species, these 35 detections may be incorrectly interpreted as false-positive detections, as demonstrated by the tendency to overestimate the false-positive encounter rate and underestimate occupancy in the simulation results. In contrast, the Bernoulli detection model, which ignores data on counts of detections, outperformed the Poisson detection model in many scenarios. The Bernoulli detection model primarily failed when a rare species coexisted with a common species, which may lead to p_{10} being higher than p_{11}. Furthermore, classification success is often poor for rare events (Sun et al., 2009). In these cases, the standard approach of constraining $p_{11} < p_{10}$ (Royle & Link, 2006) ensured biased results for rare species. In the context of conservation, surveys often target rare species, so this weakness in the Bernoulli detection model may arise in practice.

The N-mixture model produced plausible estimates from the Fort Carson and Piñon Canyon field data, while the two occupancy models did not. For example, estimating an occupancy probability of 0.00 at sites with detections, as both occupancy models did, is implausible because we expect a detected species to have a non-zero probability of presence. In particular, we strongly discount the Poisson detection model estimate that $\widetilde{z}_i = 0$ for *M. ciliolabrum* at Site 1 despite recording 16.0 bat passes per detector-night. Estimates of $\widetilde{z}_i = 0$ appear to arise from the propensity of the Poisson detection model to regard detections as false positives, as discussed above. In addition to lower overall occupancy, the Poisson detection model indicated that the two bat species rarely occupied the same site. If true, this would imply strong competitive exclusion or niche partitioning between the species, but we discount this result in favour of the N-mixture model finding that the species often co-occur.

While the data collection procedures we envision (deployment of ARUs, and classification via automatic algorithms) are relatively economical, N-mixture models can be sensitive to departures from model assumptions (Link et al., 2018). Specifically, we assume that abundance and detection conform to the distributions outlined in the Methods (negative binomial abundance, Poisson encounter rates and no additional sources of misclassification). Effectively, these assumptions are the 'cost' that must be paid to avoid manual review of data. Conventional N-mixture models are sensitive to assumption violations such as unmodelled temporal variation in abundance or detection (Barker et al., 2018; Duarte et al., 2018; Nakashima, 2020). While our proposed model performed well in initial testing, significant heterogeneity in encounter rates could induce biased occupancy estimates. If such heterogeneity-induced bias does occur, it may nonetheless be difficult to detect, which can mislead investigators (Link et al., 2018). We suspect that linking the abundance of two species increased robustness

because each species provided a check on the other species (Chambert, Grant, et al., 2018). However, further testing of assumption violations is warranted, such as performance with unmodelled variation in abundance or detection, or additional sources of false-positive detections (e.g. additional species or environmental noise). In addition, we think that some robustness to model misspecification could be achieved by incorporating validation data from a subset of sites, as we proposed in the Methods, although collecting these data would impose costs as well. In a sense, the validation-free approach we used here is a type of unsupervised learning classification model, while incorporating validation data would be an example of a supervised classification model. Such validation information would also provide direct information about encounter parameters for the species.

The main problem we detected in our approach was some bias when the encounter rate was high (Figure 2). The observed count is the product of abundance and encounter rate, and the model had difficulty separately identifying these two elements, with a tendency to overestimate abundance, especially for high encounter rates. Surveying for limited time periods may help reduce the encounter rate and therefore bias. For example, using just 1 hr of data per night would have reduced the encounter rate in our field application. Alternatively, one could partition long survey periods into bins, such as 1-hour bins and analyse the number of bins with detections, rather than the total number of detections.

We suggest that the proposed two-species false-positive N-mixture model could be widely useful because misclassifications and detection heterogeneity are common in fish and wildlife surveys (Alldredge et al., 2008; Boswell et al., 2008; Miller et al., 2012). ARUs, such as camera traps and acoustic recording devices have grown in popularity due to their ability to collect copious data, but such large datasets also increase the number of misclassifications, even when the misclassification rate is low (Gibb et al., 2019). Although our model performed well in an initial set of simulations, further testing of assumption violations is important to understand the performance of our model with field data. We applied our model to an acoustic field survey of bats, but we believe our model could be applied to a variety of taxa and ARU types. Given the increased use of ARUs for wildlife surveys, the importance of working with counts of unmarked animals and accounting for abundance-induced heterogeneity and misclassifications is likely to grow.

AUTHORS' CONTRIBUTIONS

M.J.C. and J.A.R. developed the model, wrote the code and conducted the analyses; R.J.M. designed and executed the bat surveys and identified echolocation calls. M.J.C. led the writing of the manuscript.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

PEER REVIEW

DATA AVAILABILITY STATEMENT

ORCID

Matthew J. Clement https://orcid.org/0000-0003-4231-7949
J. Andrew Royle https://orcid.org/0000-0003-3135-2167

REFERENCES

Alldredge, M. W., Pacifici, K., Simons, T. R., & Pollock, K. H. (2008). A novel field evaluation of the effectiveness of distance and independent observer sampling to estimate aural avian detection probabilities. *Journal of Applied Ecology, 45*(5), 1349–1356.

Baerwald, E. F., & Barclay, R. M. (2009). Geographic variation in activity and fatality of migratory bats at wind energy facilities. *Journal of Mammalogy, 90*(6), 1341–1349.

Barclay, R. M. (1999). Bats are not birds—A cautionary note on using echolocation calls to identify bats: A comment. *Journal of Mammalogy, 80*(1), 290–296.

Barker, R. J., Schofield, M. R., Link, W. A., & Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics, 74*(1), 369–377.

Beilke, E. A., Blakey, R. V., & O'Keefe, J. M. (2021). Bats partition activity in space and time in a large, heterogeneous landscape. *Ecology and Evolution, 11*(11), 6513–6526.

Bittle, M., & Duncan, A. (2013). A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. *Proceedings of Acoustics: Science, Technology and Amenity, 2013*, 1–8.

Boswell, K. M., Wilson, M. P., & Cowan, J. H., Jr. (2008). A semiautomated approach to estimating fish size, abundance, and behavior from dual-frequency identification sonar (DIDSON) data. *North American Journal of Fisheries Management, 28*(3), 799–807.

Campbell, M. D., Pollack, A. G., Gledhill, C. T., Switzer, T. S., & DeVries, D. A. (2015). Comparison of relative abundance indices calculated from two methods of generating video count data. *Fisheries Research, 170*, 125–133.

Chambert, T., Grant, E. H. C., Miller, D. A., Nichols, J. D., Mulder, K. P., & Brand, A. B. (2018). Two-species occupancy modelling accounting for species misidentification and non-detection. *Methods in Ecology and Evolution, 9*(6), 1468–1477.

Chambert, T., Miller, D. A., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology, 96*(2), 332–339.

Chambert, T., Waddle, J. H., Miller, D. A., Walls, S. C., & Nichols, J. D. (2018). A new framework for analysing automated acoustic species detection data: Occupancy estimation and optimization of recordings post-processing. *Methods in Ecology and Evolution, 9*(3), 560–570.

Chapman, S. S., Griffith, G. E., Omernik, J. M., Price, A. B., Freeouf, J., & Schrupp, D. L. (2006). *Ecoregions of Colorado (color poster with 1:1,200,000 scale map, text, tables, and photographs)*. Geological Survey.

Clare, J. D., Townsend, P. A., & Zuckerberg, B. (2021). Generalized model-based solutions to false-positive error in species detection/nondetection data. *Ecology, 102*(2), e03241. https://doi.org/10.1002/ecy.3241

Clement, M. J. (2016). Designing occupancy studies when false-positive detections occur. *Methods in Ecology and Evolution*, *7*(12), 1538–1547.

Clement, M. J., Murray, K. L., Solick, D. I., & Gruver, J. C. (2014). The effect of call libraries and acoustic filters on the identification of bat echolocation. *Ecology and Evolution*, *4*(17), 3482–3493.

Clement, M. J., Rodhouse, T. J., Ormsbee, P. C., Szewczak, J. M., & Nichols, J. D. (2014). Accounting for false-positive acoustic detections of bats using occupancy models. *Journal of Applied Ecology*, *51*(5), 1460–1467.

Clement, M. J., Royle, J. A. & Mixan, R. J. (2022). Fort Carson and Piñon Canyon bat acoustic survey summary data and computer code. https://doi.org/10.5281/zenodo.6496517

Duarte, A., Adams, M. J., & Peterson, J. T. (2018). Fitting N-mixture models to count data with unmodeled heterogeneity: Bias, diagnostics, and alternative approaches. *Ecological Modelling*, *374*, 51–59.

Fritsch, G., & Bruckner, A. (2014). Operator bias in software-aided bat call identification. *Ecology and Evolution*, *4*(13), 2703–2713.

Gannon, W. L., Sherwin, R. E., & Haymond, S. (2003). On the importance of articulating assumptions when conducting acoustic studies of habitat use by bats. *Wildlife Society Bulletin*, *31*(1), 45.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, *10*(2), 169–185.

Hayes, J. P. (1997). Temporal variation in activity of bats and the design of echolocation-monitoring studies. *Journal of Mammalogy*, *78*(2), 514–524.

Kellner K. (2019). jagsUI: A wrapper around 'rjags' to streamline 'JAGS' analyses. R package version 1.5.1. Retrieved from https://CRAN.R-project.org/package=jagsUI

Kéry, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and static models.* Academic Press.

Klingbeil, B. T., & Willig, M. R. (2015). Bird biodiversity assessments in temperate forest: The value of point count versus acoustic monitoring protocols. *PeerJ*, *3*, e973. https://doi.org/10.7717/peerj.973

Link, W. A., Schofield, M. R., Barker, R. J., & Sauer, J. R. (2018). On the robustness of N-mixture models. *Ecology*, *99*(7), 1547–1551.

Luczkovich, J. J., Mann, D. A., & Rountree, R. A. (2008). Passive acoustics as a tool in fisheries science. *Transactions of the American Fisheries Society*, *137*(2), 533–541.

Miller, B. W. (2001). A method for determining relative activity of free flying bats using a new activity index for acoustic monitoring. *Acta Chiropterologica*, *3*(1), 93–105.

Miller, D. A., Bailey, L. L., Grant, E. H. C., McClintock, B. T., Weir, L. A., & Simons, T. R. (2015). Performance of species occurrence estimators when basic assumptions are not met: A test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, *6*(5), 557–565.

Miller, D. A., Nichols, J. D., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, *92*(7), 1422–1428.

Miller, D. A., Weir, L. A., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Simons, T. R. (2012). Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecological Applications*, *22*(5), 1665–1674.

Mixan, R. J., Scobie, E. H., Diamond, J. M., & Ingraldi, M. (2020). Bat surveys at Fort Carson and Piñon canyon maneuver site, Arizona Game and Fish Department, Phoenix.

Nakashima, Y. (2020). Potentiality and limitations of N-mixture and Royle-Nichols models to estimate animal abundance based on noninstantaneous point surveys. *Population Ecology*, *62*(1), 151–157.

O'Connell, A. F., Nichols, J. D., & Karanth, K. U. (2010). *Camera traps in animal ecology: Methods and analyses.* Springer.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing.* ISSN 1609-395X.

R Core Team (2021). R: A language and environment for statistical computing. Version 4.04. R foundation for statistical computing. Retrieved from https://www.R-project.org/

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, *60*(1), 108–115.

Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, *87*(4), 835–841.

Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence–absence data or point counts. *Ecology*, *84*(3), 777–790.

Russo, D., Ancillotto, L., & Jones, G. (2018). Bats are still not birds in the digital era: Echolocation call variation and why it matters for bat species identification. *Canadian Journal of Zoology*, *96*(2), 63–78.

Russo, D., & Voigt, C. C. (2016). The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for a sound analysis. *Ecological Indicators*, *66*, 598–602.

Sherwin, R. E., Gannon, W. L., & Haymond, S. (2000). The efficacy of acoustic techniques to infer differential use of habitat by bats. *Acta Chiropterologica*, *2*(2), 145–153.

Shonfield, J., & Bayne, E. (2017). Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology*, *12*(1), art14.

Stanley, T. R., & Royle, J. A. (2005). Estimating site occupancy and abundance using indirect detection indices. *The Journal of Wildlife Management*, *69*(3), 874–883.

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(4), 687–719.

Thomas, D. W., Bell, G. P., & Fenton, M. B. (1987). Variation in echolocation call frequencies recorded from north American vespertilionid bats: A cautionary note. *Journal of Mammalogy*, *68*(4), 842–847.

Van Parijs, S. M., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., & Van Opzeeland, I. C. (2009). Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. *Marine Ecology Progress Series*, *395*, 21–36.

Wright, W. J., Irvine, K. M., Almberg, E. S., & Litt, A. R. (2020). Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity. *Methods in Ecology and Evolution*, *11*(1), 71–81.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.